University of Glasgow

SAVING THE CENTENARY'S DIGITAL HERITAGE

RECOMMENDATIONS FOR DIGITAL SUSTAINABILITY OF FWW COMMUNITY COMMEMORATION ACTIVITIES

Agiatis Benardou
Lorna Hughes
Leo Konstantelos

University of Glasgow, 2018

© 2018 by the University of Glasgow. This work is licensed under the Creative Commons Attribution-ShareAlike 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by-sa/4.0/

This document may be freely reviewed, copied and redistributed in any medium or format, in part or in whole. The contents may be remixed, transformed and built upon, provided that due credit is given to this document and its authors. In all cases, a link to the CC licence must be provided, with an indication of any changes made; and derivatives must be distributed under the same CC licence as the original document.

Acknowledgements

This version of the document was produced partly with funding from the Arts & Humanities Research Council and the Heritage Lottery Fund.

Contribution to the development of this document was made by the Digital Preservation Coalition (DPC), a not-for-profit company that enables organisations to deliver resilient long-term access to digital content and services, helping them to derive enduring value from digital assets and raising awareness of the strategic, cultural and technological challenges they face. For more information, visit https://www.dpconline.org/

Authors

Agiatis Benardou is Research Associate at the University of Glasgow; and Senior Research Associate at the Digital Curation Unit Athena.

Lorna Hughes is Professor of Digital Humanities, and Head of Information Studies at the University of Glasgow.

Leo Konstantelos is Lecturer in Information Studies, and Research Associate at the University of Glasgow.

Contact us

Information Studies
11 University Gardens
University of Glasgow
Glasgow, G12 8QQ
Scotland, UK.

Web: http://www.gla.ac.uk/subjects/informationstudies
Twitter: https://twitter.com/uofginfostudies

ABOUT THIS GUIDE

The Centenary of the First World War has seen a 'digital big bang': the creation of a huge number of digital collections, resources, and publications, much of it created by community projects.

This guide has been developed to advise the creators of these projects on the long term digital preservation and sustainability of these items, to make sure there will be a 'digital legacy' of the Centenary. This is something we are concerned about: we don't just want to preserve the financial investment in creating these digital resources, but the effort that went into them should be respected, and their value for future scholarship and teaching should be protected.

The guide provides useful brief summary of the principles of sustaining digitised content. It is provided as an Annex to a report on preserving outputs produced as part of your project, that have academic and historic significance, including digitised content, administrative content, as well as digital publishing platforms.

01 Who are you saving your digital outputs for

To ensure sustainability of the Centenary's Digital Heritage, community engagement should be sought throughout the full lifecycle of initiatives

WHY IT IS IMPORTANT

Communities of practice and interest are the main recipients of your digital outputs, and the ones to reuse and thus sustain them.

PUTTING IT INTO PRACTICE

- Identify who the audience for your digital outputs is (e.g. (teachers, students, researchers, general public).
- Analyse your audience's needs for access and sustainability of your digital outputs.
- Create a community, by consulting with your audience about decisions that might affect their access to your digital outputs.

HOW TO IDENTIFY AND ANALYSE YOUR AUDIENCE

1

Identify target audience

Profile target audience

use to design and deliver community engagement activities

Brainstorm and list all target audiences that are interested in your digital outputs for work or personal use

Use **existing evidence**, e.g. from traffic data you collect via your website; or from surveys and events that you have previously organised

Collaborate with teams and individuals in your organisation for richer and deeper insights into the potential audiences

Try to include people who have **direct experience** working with your communities of practice and interest

Identify **target audience characteristics**

Socio-demographic	Geographic	Psychographic	Communication channels	Interests and practices
• Gender • Age range • Education • Employment • Occupation • Language	• Region • City • Location of access to digital outputs	• Needs for using digital outputs • Attitudes and opinions	• Frequent v. occasional use • Reasons for accessing digital outputs • Comms media to reach audience	• Current activities • Current knowledge relating to digital outputs • Ability and know-how to access digital outputs
Sources: • Census data • Surveys and interviews • Data from similar projects	Sources: • Census data • Local government data • Surveys and interviews	Sources: • Market research • Project research • Media reports	Sources: • Community media • Printed press • Radio and TV • Social media	Sources: • Community events • Market research • Project research

02 What have you produced

To manage and maintain your digital outputs in the long-term, you need to know what you have generated throughout the project that is of academic and historic significance

WHY IT IS IMPORTANT

Without an accurate account of the digital outputs that you have produced, it will not be possible to decide *why* you should keep them. Your digital outputs may be lost, deleted or not available when needed. Undocumented digital outputs are at risk of becoming invisible – and therefore inaccessible - to you and your community.

PUTTING IT INTO PRACTICE

- Audit the digital outputs you have produced, by following a process to identify, locate, describe and assess them.
- Create a detailed inventory, which shows what you have, where it is stored, if it's secure and in good condition, who is responsible for managing it, and that you have the correct permissions to use it.

TYPES OF DIGITAL OUTPUTS WITH EXAMPLES

2

Use the following as a guide to organise your digital outputs by **type**. The examples of common file formats and content will help you decide about the type of an output, if you are uncertain

TEXT & DOCUMENTS

COMMON FILE FORMATS

Microsoft office documents
- Text – DOC, DOCX
- Spreadsheets – XLS, XLSX
- Presentations – PPT, PPTX

Open Office Documents
- Text – ODT
- Spreadsheets – ODS
- Presentations – ODP

Other
- Plain text – TXT
- Rich text format – RTF
- Optical Character Recognition

Example of OCR text from Cymru1914

IMAGES & PHOTOS

COMMON FILE FORMATS

Photographs
- Adobe Photoshop Document – PSD
- Portable Network Graphics – PNG
- Tagged-Image File Format – TIFF

Graphics and logos
- Graphics Interchange Format – GIF
- Joint Photographic Experts Group – JPG, JPEG
- Windows Bitmap - BMP

Vector graphics
- Encapsulated PostScript – EPS
- Adobe Illustrator – AI
- Scalable Vector Graphics – SVG

Example of digitised images

AUDIO & VIDEO

COMMON FILE FORMATS

Audio
- Pulse-Code Modulation – PCM
- Waveform Audio – WAV
- Audio Interchange – AIFF
- MPEG-1 Audio – MP3
- Advanced Audio Coding – AAC
- Windows Media Audio - WMA

Video
- Audio Video Interleave – AVI
- Windows Media Video – WMV
- Apple QuickTime Movie – MOV
- Moving Pictures Expert Group – MP4

Example of digital audio

3D OBJECTS

COMMON FILE FORMATS

Models and rendering
- Stereolithography – STL
- Wavefront 3D Object File – OBJ
- Autodesk FBX – FBX
- 3D Studio Scene – 3DS
- IGES File – IGES
- STEP 3D Model – STEP
- Xara3D Project – X3D
- Virtual Reality Modeling Language – VRML

Example of 3D representation

03 How to save in sustainable file formats

File formats – commonly identified by looking at the file extension – define the structure and type of your digital outputs, so that machines understand how to access and render them properly

WHY IT IS IMPORTANT

File formats are prone to **obsolescence with the passing of time**. Formats evolve as users and developers identify and incorporate new functionality. New formats, or versions of formats, may introduce file format obsolescence as newer generations of software phase out support for older formats.
When software does not provide for backwards compatibility with older file formats, **digital outputs may become unusable**. (DPC Handbook)

PUTTING IT INTO PRACTICE

- Convert digital outputs in formats at risk of obsolescence into sustainable formats.
- Use this guide to identify sustainable formats for different types of digital outputs.
- This process requires some degree of technical expertise.

CHOOSING SUSTAINABLE FILE FORMATS

3

Use the following as a guide to convert your current digital outputs into sustainable file formats

You don't need to convert digital outputs that are already saved in a sustainable format

Convert…	…into this sustainable file format
Text documents	• Open document Text Format (ODT) • Portable Document Format/Archival (PDF/A-2)
Photographs and graphics	• Portable Network Graphics (PNG) • Tagged Image File Format (TIFF) • JPEG 2000 (JP2)
Vector graphics and illustrations	• Scalable Vector Graphics (SVG)
Audio files	• Audio Interchange (AIFF) • WAVeform Audio (WAV)
Video files	• Motion JPEG 2000 (MJP2 or MJ2) • MPEG-4
Presentations and slides	• OpenDocument Presentation Format (ODP) • Portable Document Format/Archival (PDF/A-2)

04 How to choose sustainable computer storage

Store your digital outputs in computer environments that will facilitate current access and provide future-proof solutions

WHY IT IS IMPORTANT

Much like file formats, data storage is also prone to obsolescence and physical threats. Removable media, such as CDs and DVDs, can be easily lost or damaged. Hard disk drives can malfunction or break. Legacy storage media, such as floppy disks, may not be supported by contemporary computer environments.

PUTTING IT INTO PRACTICE

- Save and organise all your digital outputs in one central location – make sure that you include data stored in removable media.
- Create copies of all your digital outputs that are held in at least two different locations.
- Establish backup procedures for your digital outputs, either using your organisation's IT solutions or through external providers.

CHOOSING SUSTAINABLE STORAGE

4

Use the following as a guide to select storage solutions for your digital outputs that will ensure against data loss

DIGITAL OUTPUTS SAVED IN REMOVABLE AND PORTABLE MEDIA

Some of your digital outputs may be stored in **Removable storage media**, which include:

- CDs and DVDs
- Blue Ray disks
- USB memory sticks and flash drives
- memory cards
- as well as legacy media, such as floppy disks.

You may also have digital outputs stored in **Portable media**, such as:

- Laptops
- Smartphones
- Tablets

These are not considered stable storage media, and are therefore not suitable for sustainability

Transfer all digital outputs in removable and portable media to fixed media in the first place, e.g. a Desktop PC

CHOOSING SUSTAINABLE STORAGE

4

Use the following as a guide to select storage solutions for your digital outputs that will ensure against data loss

DIGITAL OUTPUTS SAVED IN FIXED MEDIA

Stand-alone computers, such as Desktop PCs and Apple Macs, and **external hard disk drives (HDD)** are a safer environment to store digital outputs than removable and portable media, **provided that:**

- They are regularly checked for **hard disk drive integrity**; and against **computer viruses**.
- Access is regulated and allowed only to **authorised users**, who will not accidentally delete or damage files.
- The location in which they are kept provides **protection against physical damage and environmental hazards** (especially water ingress and electromagnetic fields that can compromise the operation of mechanical and electronic parts).
- **Multiple copies** of the digital outputs are created and stored in **different physical locations**.
- Regular **backups** are taken (see next page).

Gather all digital outputs in one stand-alone computer. Use Section 5 to organise and document your files, then create multiple copies in external HDD and backups.

CHOOSING SUSTAINABLE STORAGE

4

SOLUTIONS FOR CREATING BACKUPS

Keeping all your digital outputs in one stand-alone computer is only the first step towards a safe and sustainable storage environment. If the hard disk drive on your stand-alone computer fails, it can be expensive – and oftentimes impossible – to recover lost files.

Most modern Operating Systems, such as Microsoft Windows and MacOS, include software for scheduling, managing and running regular backups.

Use one or a combination of the following backup storage options to ensure that your digital outputs remain safe.

OPTION 1: MUTIPLE EXTERNAL HARD DRIVES (EHD)

These are exactly like the hard drives inside stand-alone computers, except that they are packed in a separate case and connect via USB. EHD can also fail, and are prone to obsolescence.

For optimum safety, make sure that you:

- store them in a dry, secure location.
- check regularly that they still function properly.
- create at least two copies of your digital outputs on two separate EHDs that are stored in different locations.

OPTION 2: NETWORK ATTACHED STORAGE (NAS)

A NAS is a storage device connected to a network that allows storage and retrieval of data from a central location for authorised users. NAS are a faster and more secure solution than managing multiple EHDs; and are focused on creating file backups in simple enough ways to not necessitate IT expertise.

There are many companies that provide NAS solutions, often arranging to set up the devices in your premises and train staff to use them.

OPTION 3: THIRD-PARTY CLOUD BACKUP SERVICES

These involve paying a fee to a third-party service provider, in order to send copies of your digital outputs over a network. The fee includes the provider's costs for maintaining your backups; and depends largely on the amount of storage you need.

If the funds to pay third-party fees are available, cloud backup services may be a cost-effective option, as they reduce the need for IT expertise. Some providers even offer free cloud storage up to a certain limit of data volume.

05 What to document for the present and future

Organise, document and contextualise your digital outputs, so that they can be understood, interpreted and discovered now and into the future

WHY IT IS IMPORTANT

Digital sustainability relies as much on file formats and storage, as on simple information management practices.

Organising your digital outputs in a logical file structure allows for easier access and discoverability – by you and others.
Documenting your digital outputs means that you or others can more easily understand and interpret the content and context of your files.

PUTTING IT INTO PRACTICE

- Establish file naming conventions that will help you consistently name files and folders.
- Use methods for documenting the content and context of your digital outputs.
- Create a system for organising your files into folders.

ESTABLISHING INFORMATION MANAGEMENT PRACTICES

5

Use the following as a guide to organise, document and contextualise your digital outputs

FILE & FOLDER NAMING

Agreeing on a naming convention will help to provide consistency, which will make it easier to find and correctly identify your digital outputs.

There is no prescribed way to create file name conventions, as long as:

- it resonates with the way that your team and organisation operate;
- it is fully documented, and the documentation available to your peers;
- it is applied consistently and accurately to all your digital outputs.

When creating a naming convention, consider the following:

- Terminology, language and abbreviations
  e.g. "WW1" or "FWW"
- Order and sequence
  e.g. Date-Project code-Description
- Dates and numbers
  e.g. YYYY-MM-DD; files numbered as 001-00x
- Punctuation and symbols
  e.g. use of capital letters, hyphens and underscores

FOLDER ORGANISATION

Organising your digital outputs into a logical folder structure will help you access your files more easily and quickly, avoid duplication, and provide context and meaning for current and future use.

When creating a folder structure to organise your digital outputs, consider the following:

- Group similar files together in folders and sub-folders
  store digital outputs that belong to the same themes and sub-themes together. E.g.:

```
"Images" folder
 ├──"Digitised collections" folder
 │   ├────"Life in the trenches" folder
 │   ├────"Women at the war" folder
```

- Structure your folder groups hierarchically
  create first-level folders for the broader topics, then create more specific folders within them.
- Use the naming convention
  name your folders using the naming convention that you have agreed on. As for files, be consistent and accurate when following the convention.

DOCUMENTATION

Documenting your digital outputs with descriptive information about their content and context will help ensure that they can be understood by others now and in the future.

When documenting your digital outputs, consider what information you need to provide so that someone completely unfamiliar with your project or work can interpret their meaning in five, ten or fifty-years time.

There are several ways to document your digital outputs, including:

- Embedded documentation
  certain software, such as Microsoft Office, allows users to embed documentation into the file itself.
- Supporting documentation in separate files
  create simple text files that you save in each folder and/or sub-folder, and that provide information about the contents of that folder. See this useful guide to get started.
- Use metadata
  for advanced users and owners of repository software.

FURTHER INFORMATION

Helpful resources

Digital Preservation Coalition (DPC) - https://www.dpconline.org/handbook

The National Archives - http://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/

Collections Trust - https://collectionstrust.org.uk/digital-isnt-different/digital-preservation/

Library of Congress (LOC) - http://www.digitalpreservation.gov/

Digital Curation Centre (DCC) – http://www.dcc.ac.uk/resources/how-guides

University of Glasgow

SAVING THE CENTENARY'S DIGITAL HERITAGE

RECOMMENDATIONS FOR DIGITAL SUSTAINABILITY OF FWW COMMUNITY COMMEMORATION ACTIVITIES

by Agiatis Benardou, Lorna Hughes and Leo Konstantelos

© 2018 by the University of Glasgow. This work is licensed under the Creative Commons Attribution-ShareAlike 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by-sa/4.0/